Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Three-centre C—H···O hydrogen bonds in the DNA minor groove: analysis of oligonucleotide crystal structures

Anirban Ghosh and Manju Bansal*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Correspondence e-mail: mb@mbu.iisc.ernet.in

Received 8 July 1999
Accepted 8 October 1999

AA·TT and GA·TC dinucleotide steps in B-DNA-type oligomeric crystal structures and in protein-bound DNA fragments (solved using data with resolution <2.6 Å) show very small variations in their local dinucleotide geometries. A detailed analysis of these crystal structures reveals that in AA·TT and GA·TC steps the electropositive C2—H2 group of adenine is in very close proximity to the keto O atoms of both the pyrimidine bases in the antiparallel strand of the duplex structure, suggesting the possibility of intra-base pair as well as cross-strand inter-base pair C—H···O hydrogen bonds in the DNA minor groove. The C2—H2···O2 hydrogen bonds in the A·T base pairs could be a natural consequence of Watson–Crick pairing. However, the cross-strand interactions between the bases at the 3′-end of the AA·TT and GA·TC steps obviously arise owing to specific local geometry of these steps, since a majority of the H2···O2 distances in both data sets are considerably shorter than their values in the uniform fibre model (3.3 Å) and many are even smaller than the sum of the van der Waals radii. The analysis suggests that in addition to already documented features such as the large propeller twist of A·T base pairs and the hydration of the minor groove, these C2—H2···O2 cross-strand interactions may also play a role in the narrowing of the minor groove in A-tract regions of DNA and help explain the high structural rigidity and stability observed for poly(dA)·poly(dT).

1. Introduction

One of the most striking and recurring sequence-dependent features observed in DNA oligomer crystal structures is the narrowing of the minor groove in regions containing tracts of adenine–thymine base pairs (A-tracts). Recent high-resolution crystal structures of the well studied Dickerson dodecamer sequence d(CGCGAATTCGCG) have confirmed that the AT-rich region is straight and the minor groove is narrow (Shui, McFail-Isom *et al.*, 1998; Shui, Sines *et al.*, 1998; Tereshko *et al.*, 1999). These characteristic features of A-tracts have generally been attributed to the presence of a spine of hydration in the minor groove (Drew & Dickerson, 1981) or to the occurrence of high propeller twist in AT base pairs, leading to favourable cross-strand interactions between the 6-amino group of adenine and the 4-keto O atom of thymine in the major groove of an A-tract (Nelson *et al.*, 1987; DiGabriele & Steitz, 1993). The repertoire of major-groove cross-strand hydrogen bonds has recently been expanded to include favourable N—H···N interactions between the 6-amino groups of the two 5′-adenines in AT·AT steps and with the 4-amino group of cytosine in the inosine-containing steps AI·CT (Shatzky-Schwartz *et al.*, 1997; Luisi *et al.*, 1998).

© 1999 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Mean values of propeller twist and dinucleotide step parameters in the fibre-model structure (Chandrasekaran & Arnott, 1996), all dinucleotide steps in the oligomer (o) and protein-complexed (c) DNA crystal structures, and in the AA·TT and GA·TC steps only in the two data sets.

The corresponding standard deviation values are given below the mean values. The number of data points in each case is given in parentheses.

	Propeller (°)	Tilt (°)	Roll (°)	Twist (°)	Shift (Å)	Slide (Å)	Rise (Å)
Fibre	−14.7	−0.2	2.2	35.9	0.04	0.56	3.35
Oligomer (328)	−11.7	0.1	1.5	35.7	0.02	0.33	3.35
	6.6	2.1	5.5	5.2	0.47	0.67	0.11
Complex (373)	−10.7	0.0	2.8	34.1	0.00	−0.12	3.34
	6.0	2.3	4.9	4.7	0.54	0.65	0.12
AA·TT o (53)	−16.2	−0.1	0.9	35.2	−0.02	−0.18	3.26
	4.5	1.7	3.9	3.5	0.22	0.34	0.07
AA·TT c (45)	−14.3	0.9	−0.3	36.1	0.09	−0.04	3.26
	7.3	2.3	3.9	3.3	0.45	0.45	0.16
GA·TC o (40)	−12.9	−0.1	0.4	38.2	−0.01	−0.02	3.29
	4.5	1.9	3.7	3.2	0.37	0.38	0.08
GA·TC c (39)	−11.6	−0.3	2.8	35.7	0.05	0.06	3.30
	6.2	2.8	4.3	4.1	0.53	0.67	0.11

We have recently analysed all possible cross-strand interactions, in both the major and minor grooves of duplex DNA, for the ten dinucleotide sequences in DNA oligomer and nucleic acid–regulatory protein complex structures (Bansal & Ghosh, 1999). In addition to N···O and N···N separation, which has generally been used to assign a hydrogen bond, the H···O or H···N distances and N—H···O/N angles, which would be a true chemical indicator of hydrogen-bond formation, have also been examined. It is found that very few cross-strand N—H···O interactions in either the major or minor groove can actually be classified as hydrogen bonds by these composite criteria, while a slightly larger proportion of N—H···N interactions in the major groove of AT·AT, CG·CG and AG·CT steps meet the hydrogen-bonding criteria (Bansal & Ghosh, 1999). We also examined the C2···O2 and H2···O2 distances in these structures to identify possible C—H···O hydrogen bonds in the Watson–Crick base pairs as well as in the minor groove of AA·TT and GA·TC steps. Such hydrogen bonds have been reported to occur in the structures of organic molecules (Taylor & Kennard, 1982; Desiraju, 1991, 1996; Steiner & Saenger, 1993), nucleic acids (Leonard *et al.*, 1995; Wahl & Sundaralingam, 1997; Auffinger & Westhof, 1998), proteins (Derewenda *et al.*, 1995; Bella & Berman, 1996; Fabiola *et al.*, 1997; Chakrabarti & Chakrabarti, 1998) and protein–DNA complexes (Mandel-Gutfreund *et al.*, 1998). It was found that the AA·TT and GA·TC dinucleotide steps do seem to prefer geometries which simultaneously favour C—H···O interactions between the A·T Watson–Crick base pairs and cross-strand C—H···O interactions (hydrogen bonds) in the minor groove between the bases at the 3′-ends of the AA·TT and GA·TC steps.

2. Data and methods

The data available in the December 1998 release of the Brookhaven Protein Data Bank (PDB) was used as the source

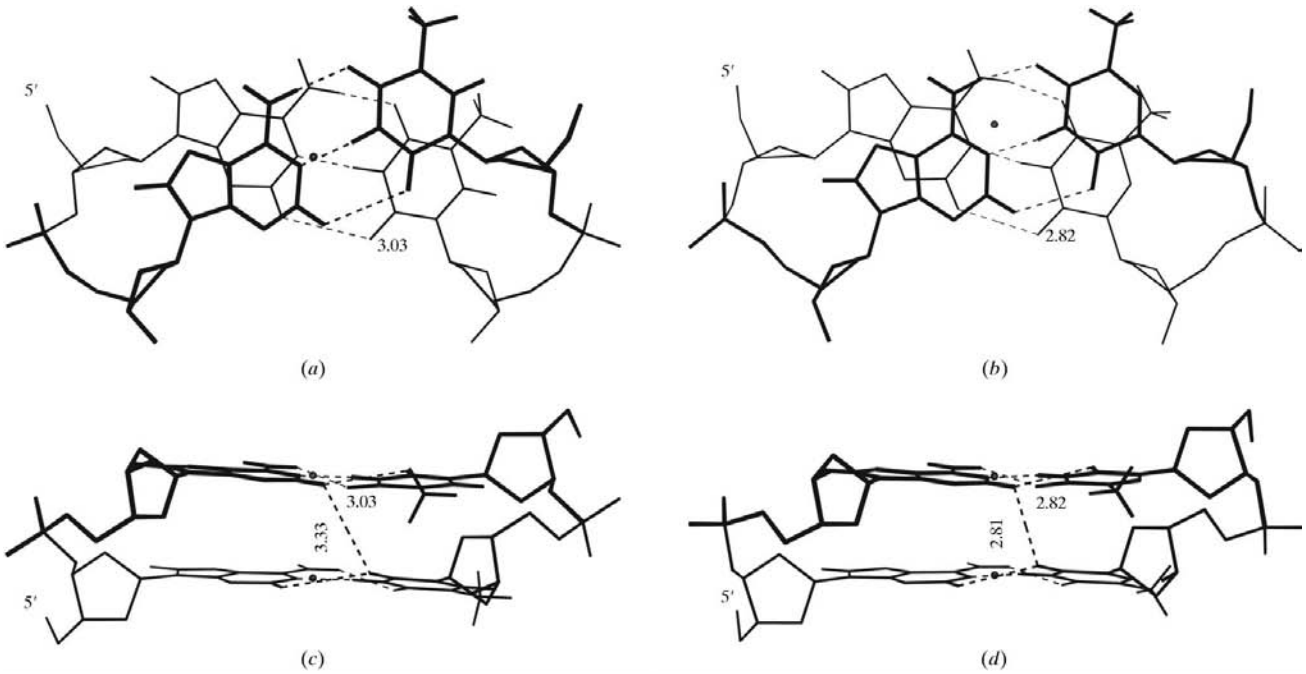

Figure 1
Projection down the helix axis showing two Watson–Crick-paired A·T base pairs in an AA·TT step generated with (*a*) the dinucleotide geometry as in the fibre-model structure and (*b*) the average geometry from the oligomer crystal structure data set (with parameters listed in Table 1). H2···O2 distances corresponding to potential (Ade)C2—H2···O2(Thy) hydrogen bonds in the A·T base pairs have been indicated. A view into the minor groove of the same AA·TT steps is shown in (*c*) and (*d*). The cross-strand H2···O2 distances in the minor groove between the 3′-end adenine in strand 1 and 3′-end thymine in strand 2 are indicated, along with the H2···O2 distances in the Watson–Crick base pairs.

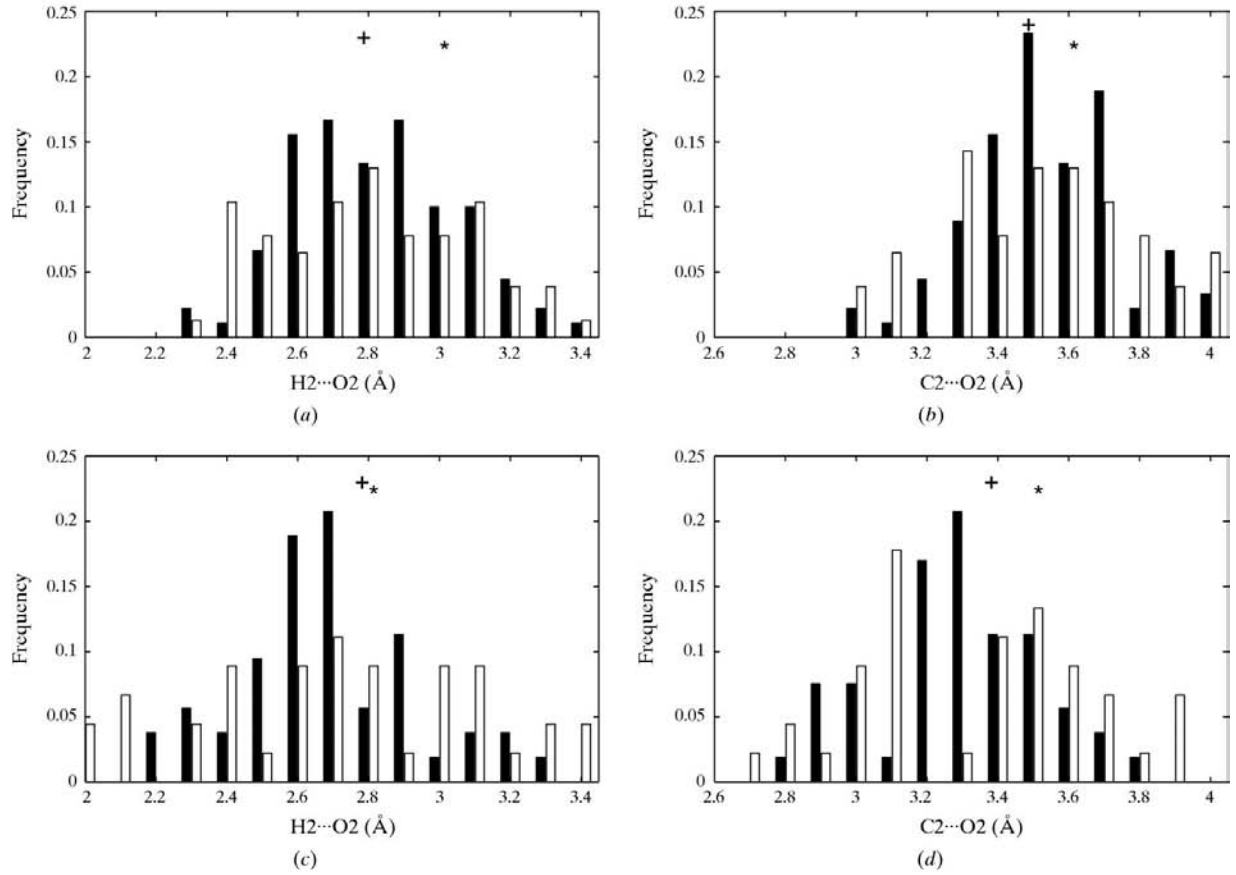

Figure 2
Histograms showing the frequency of occurrence of H2···O2 (*a* and *c*) and C2···O2 (*b* and *d*) cross-strand distances in the ranges 2.0–3.4 Å and 2.6–4.0 Å, respectively, in the Watson–Crick A·T base pairs (*a* and *b*) and minor grooves of AA·TT (*c* and *d*) steps. The data for B-DNA type oligomer crystal structures are shown as solid bars and the corresponding values for protein-bound DNA structures are represented by hollow bars. The frequency of occurrence has been normalized so that data sets of different sizes can be compared. There are 53 AA·TT steps in the oligomer data set and 45 in the DNA–protein data set; their mean values are indicated by a plus sign and an asterisk, respectively, in each histogram.

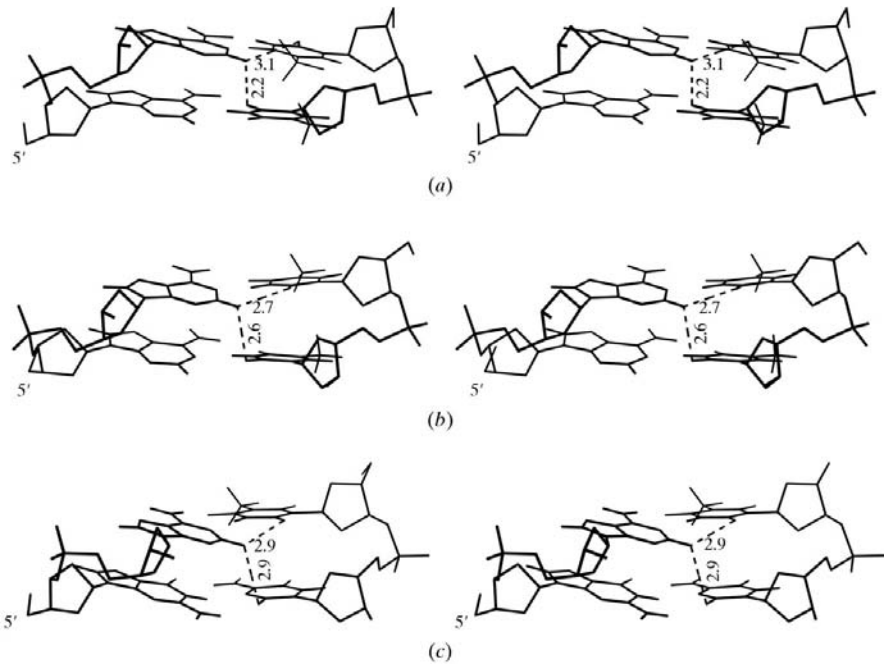

Figure 3
Stereo diagram showing representative examples of AA·TT and GA·TC steps in crystal structures of oligomeric DNA. (*a*) A6A7·T6T7 step in the duplex structure with the sequence 5′-CGCAAAAAACGC-3′ in strand 1 (1d98; Nelson *et al.*, 1987) and (*b*) A5A6·T7T8 step in high-resolution (1.1 Å) DNA structure with the self-complementary sequence 5′-CGCGAATT-CGCG-3′ (PDB code 436d; Tereshko *et al.*, 1999). (*c*) G4A5·T8C9 step in the 436d structure. H2···O2 distances in the 3′-end A·T base pair and between the 3′-end adenine in strand 1 and 3′-end thymine or cytosine in strand 2 are shown in all cases.

Table 2
Mean values and standard deviation for distances and angles corresponding to C2—H2···O2 hydrogen bonds in AA·TT steps of DNA oligomer and protein-complexed structures.

Data for AA·TT steps in crystal structures with <2.0 Å resolution have been listed explicitly. Watson–Crick-type hydrogen-bonding distances have been listed in this case only for the adenine involved in the cross-strand C—H···O interaction. The base sequence for oligomer structures in the first strand are: 436d, 355d, 1bna (5′-CGCGAATTCGCG-3′), 1d49 (5′-CGATTAATCG-3′), 307d (5′-CAAAGAAAAG-3′), 158d (5′-CCAAGCTTGG-3′), 5dnb (5′-CCAACGTTGG-3′) and 1d61 (5′-CCAACITTGG-3′). The last two structures have twofold symmetry. For protein-complexed DNA structures, the sequences for the first strand are: 1hcr (5′-GTTTTTGATAAGA-3′), 1tro (5′-GTACTAGTTAACTAGTAC-3′) and 1lat (5′-TCCAGAACATGTTCTGGA-3′).

		Watson–Crick			Cross-strand		
Structure	Resolution (Å)	C2···O2 (Å)	H2···O2 (Å)	C2—H2···O2 (°)	C2···O2 (Å)	H2···O2 (Å)	C2—H2···O2 (°)
Fibre		3.71	3.03	121.2	4.05	3.33	124.9
Oligomer		3.54 (0.21)	2.82 (0.22)	123.8 (7.0)	3.43 (0.46)	2.81 (0.45)	116.5 (7.21)
Complex		3.63 (0.36)	2.96 (0.44)	120.68 (11.72)	3.47 (0.45)	2.84 (0.51)	117.3 (12.38)
436d	1.1	3.45	2.70	125.5	3.21	2.59	115.4
		3.46	2.73	123.5	3.11	2.49	115.5
355d	1.4	3.43	2.65	127.6	3.31	2.74	112.2
		3.45	2.69	125.9	3.24	2.64	114.0
1bna	1.9	3.49	2.78	122.8	3.29	2.66	116.5
		3.32	2.61	121.8	3.32	2.74	112.9
1d49	1.5	3.32	2.54	128.1	3.22	2.66	111.6
		3.71	2.94	127.7	3.54	2.87	119.9
307d	1.85	3.48	2.80	120.4	4.14	3.50	119.3
		3.60	2.92	120.1	3.80	3.29	110.1
		3.62	2.91	122.8	3.00	2.37	115.1
		3.51	2.85	119.3	2.91	2.30	113.3
		3.34	2.59	125.9	3.26	2.64	115.3
158d	1.9	3.62	2.88	125.0	5.21	4.43	130.7
		3.50	2.72	128.5	4.85	4.12	127.1
5dnb	1.4	3.40	2.61	129.4	4.53	3.84	123.2
1d61	1.3	3.38	2.64	124.1	4.51	3.90	117.7
1hcr	1.8	3.46	3.05	103.4	3.43	3.11	97.9
		3.12	2.75	99.5	3.08	2.30	126.9
		3.71	3.24	107.1	3.32	2.78	110.4
		3.29	2.45	133.0	4.10	3.32	129.0
		3.46	2.42	158.5	4.20	3.58	117.8
1tro (mol 1)	1.9	3.51	2.78	124.0	3.56	2.97	114.6
		3.62	2.96	119.7	3.54	2.82	123.3
1tro (mol 2)		3.65	2.93	123.8	3.46	2.94	109.8
		3.80	3.07	124.9	3.56	2.96	115.1
1lat	1.9	3.70	3.03	120.7	3.41	2.57	133.8
		3.80	3.09	123.7	3.53	2.73	129.7

of the crystal structure coordinates. 35 oligomeric B-DNA-type structures, including those with modified bases but excluding those with base-pair mismatches, solved using X-ray data with resolution <2.6 Å were included in the database [PDB codes: 1bna, 355d, 1d98, 1dn9, 1bdn, 1d29, 1d65, 119d, 1d89, 249d (two molecules), 271d, 194d, 4bna, 4dnb, 1d77, 265d, 270d, 1bd1, 5dnb, 1d23, 1d49, 1d56, 1d57, 1cgc, 126d, 158d, 167d, 196d, 252d, 307d, 2d25, 1d60, 1d61, 1da3, 183d]. The data set contains 53 AA·TT steps consisting of 90 A·T base pairs and 40 GA·TC dinucleotide steps with the same number of A·T and G·C base pairs.

The database of regulatory protein-bound DNA fragments consists of 20 structures (containing 24 molecules in the asymmetric units) which have been solved with a data resolution <2.6 Å [PDB codes: 1aay, 2bop, 1lmb, 1hcr, 1tro (two molecules), 1lat, 1zaa, 1lli, 1mey (two molecules), 1tup, 2dgc, 1tsr, 1trr (two molecules), 1hcq (two molecules), 1pdn, 1per, 1rpe, 1ubd, 2or1, 3cro]. This data set contains 45 AA·TT steps consisting of 77 A·T base pairs and 39 GA·TC dinucleotide steps.

H atoms have been fixed in the adenine base plane using standard bond lengths [b(C—H) = 1.08 Å] and angles (X—C—H = 120°) for all the adenines in the database. The interatomic distances between the C2 and H2 atoms of adenine on one strand of the DNA duplex and the O2 atom of thymine or cytosine (as well as the N2 of guanine) on the other strand have been calculated. We have analysed the distribution of H···O and H···N distances and the angles (donor atom)C2—H2···O2/N2(acceptor atom) in order to check which cross-strand interactions can be truly classified as being hydrogen bonds using these criteria rather than merely the donor···acceptor distances. The relevant distances in the latest B-DNA fibre-model structure (Chandrasekaran & Arnott, 1996) were also calculated for comparison. The dinucleotide step parameters (as defined by Dickerson *et al.*, 1989) were calculated using the *NUPARM* package (Bansal *et al.*, 1995).

3. Results and discussion

3.1. C—H···O hydrogen bonds in Watson–Crick-type A·T base pairs

The possibility of C—H···O hydrogen-bond formation in (Ade)C2—H2···O2(Thy/Ura) in Watson–Crick base pairs, (Ade)C8—H8···O2(Thy/Ura) in Hoogsteen base pairs, (Ade)C2—H2···O8(8-oxoGua) in an A(*anti*)·O8G(*syn*)-type base pair and (Ura)C5—H5···O4(Ura) has been suggested recently from crystal structure analysis (Leonard *et al.*, 1995; Wahl & Sundaralingam, 1997). However, no detailed analysis of such hydrogen bonds for all available DNA oligomer structures has been reported. While the C—H···O interactions in the Watson–Crick base pairs will be primarily determined by the geometry of the A·T/U base pairs and provide a third hydrogen bond to base pairs already stabilized by two conventional (N—H···O and N—H···N) hydrogen bonds, some variability can be expected owing to inter-base movement arising from propeller twisting and opening of the base pair. Fig. 1(*a*) shows a view down the helix axis of an AA·TT dinucleotide step generated with the local parameters listed in Table 1 and corresponding to the latest fibre model of B-DNA (Chandrasekaran & Arnott, 1996), while Fig. 1(*b*) corresponds to a dinucleotide generated with the average AA step parameters from the oligomer crystal structure database. The (Ade)H2···O2(Thy) distance in the Watson–Crick base

Table 3
Mean values and standard deviation for distances and angles corresponding to C2–H2···O2 hydrogen bonds in GA·TC steps of DNA oligomer and protein complexed structures.

Data for GA·TC steps in crystal structures with <2.0 Å resolution have been listed explicitly. Base sequences for structures not included in Table 2 are: 1d23 (5′-CGATCGATCG-3′), 196d (5′-CTCTCGAGAG-3′), 1d56 (5′-CGATATATCG-3′), 2bop (5′-CGACCGACGTCGGTCG-3′), 1lmb (5′-ATACCACTGGCGGTGATAT-3′).

		Watson–Crick			Cross-strand		
Structure	Resolution (Å)	C2···O2 (Å)	H2···O2 (Å)	C2—H2···O2 (°)	C2···O2 (Å)	H2···O2 (Å)	C2—H2···O2 (°)
Fibre		3.71	3.03	121.2	4.08	3.34	125.7
Oligomer		3.56 (0.18)	2.84 (0.18)	123.9 (7.13)	3.54 (0.32)	2.95 (0.34)	114.2 (7.21)
Complex		3.58 (0.29)	2.91 (0.35)	120.8 (8.08)	3.74 (0.54)	3.15 (0.54)	114.5 (9.37)
436d	1.1	3.62	2.95	119.9	3.56	2.91	117.9
		3.54	2.82	123.4	3.39	2.79	114.8
355d	1.4	3.56	2.87	121.2	3.43	2.86	112.2
		3.46	2.69	127.2	3.35	2.75	114.7
1bna	1.9	3.55	2.80	125.9	3.50	2.85	118.0
		3.52	2.92	115.2	3.17	2.50	118.3
1d49	1.5	3.59	2.86	124.5	3.52	3.00	109.6
		3.48	2.67	131.0	3.43	2.83	114.5
1d23	1.5	3.60	2.86	125.1	3.42	2.93	107.6
		3.46	2.71	125.0	3.21	2.64	112.0
		3.62	2.91	122.8	3.83	3.19	118.6
		3.50	2.74	127.1	3.41	2.79	115.7
196d	1.7	3.70	2.96	125.1	3.94	3.37	113.4
		3.62	2.86	126.3	4.62	3.88	127.1
		3.68	2.89	129.9	3.95	3.29	119.6
		3.71	2.90	131.2	3.59	3.03	112.4
1d56	1.7	3.80	3.06	125.5	3.40	2.75	117.9
		3.46	2.74	123.5	3.49	2.90	113.8
307d	1.85	3.49	2.70	129.0	3.42	2.83	113.9
2bop	1.7	3.92	3.32	116.3	4.61	3.98	118.7
		3.71	3.04	119.9	3.51	3.00	109.0
		3.71	3.04	119.8	3.51	3.00	109.1
		3.92	3.32	116.3	4.61	3.98	118.7
1lmb	1.8	3.62	2.94	120.7	3.31	2.70	114.9
1hcr	1.8	3.75	3.01	125.6	4.14	3.65	109.0
		4.15	3.45	122.8	2.87	2.21	117.0
1lat	1.9	3.75	3.04	123.6	3.49	2.74	125.7
		3.72	3.11	116.1	3.55	2.83	124.1
		3.68	3.01	119.9	3.46	2.75	122.2
		3.71	3.12	115.3	3.61	2.79	132.0

pairs in the model structures is indicated in each case; it is interesting to note that while the average value for the AA·TT step in the fibre model is 3.03 Å, in the oligomer crystal structures it is reduced to 2.82 Å, even though the average propeller twist of A·T base pairs in these steps is slightly larger (−16.2°) than in the fibre model (−14.7°). The corresponding C2···O2 distances are 3.71 and 3.54 Å, respectively. A very similar trend is observed for the GA·TC steps in the oligomer structures.

Tables 2 and 3 list the mean values of the parameters corresponding to the potential C—H···O hydrogen bonds in AA·TT and GA·TC steps in the oligomer and in protein-complexed crystal structures. The values for crystal structures with data resolution <2.0 Å are given explicitly for individual AA·TT and GA·TC steps in the structures. In addition to the structures included in the detailed statistical analysis, the data for a recently reported 1.1 Å resolution structure (436d; Tereshko *et al.*, 1999) has also been listed. It is seen that while the C—H···O angles remain close to 120°, a large number of C2···O2 and H2···O2 distances in the crystal structures (particularly those listed in Tables 2 and 3 for the high-resolution structures) are smaller than the mean values of the data set as well as the fibre-model values, indicating that close C2—H2···O2 interactions in A·T base pairs are very common and favourable. The base pairs in the protein-complexed DNA structures show slightly more variability, as is evident from the standard-deviation values of both distances and the C—H···O angle.

The histograms in Figs. 2(*a*) and 2(*b*) show the overall distribution (as a fractional frequency of occurrence) of H2···O2 and C2···O2 distances in the A·T Watson–Crick base pairs constituting the AA·TT steps in the oligomer and complex crystal structures. While the C2···O2 distances cluster about the mean values of the crystal data sets, with the maximum being between 3.4 and 3.7 Å, the H2···O2 distance shows a wider spread (ranging from 2.5 to 3.1 Å), indicating that it is more variable because of small changes in intra-base-pair parameters.

3.2. Cross-strand C—H···O hydrogen bonds

Figs. 1(*c*) and 1(*d*) show an edge-on view into the minor groove of AA·TT steps in the fibre model and the oligomer crystal structure data set (generated with the average geometry given in Table 1). It is seen that the cross-strand distance between the H2 of adenine at the 3′-end of the AA strand and the O2 of the thymine at the 3′-end of the TT strand is decreased distinctly from 3.33 Å in the fibre to 2.81 Å in the average crystal geometry, even though both steps have very similar propeller twist and step parameters, with the exception of the translation slide parameter (as seen in Table 1). The parameters corresponding to potential cross-strand C—H···O hydrogen bonds in the fibre model, the mean values in both the crystal data sets and those for structures with data resolution <2.0 Å are listed in Tables 2 and 3 for AA·TT and GA·TC steps, respectively, along with the parameters for the corresponding A·T Watson–Crick base pairs. Figs. 3(*a*) and 3(*b*) show stereo diagrams of two representative examples of AA·TT steps from the oligomer data set. The H2···O2 distances between the H2 atom of 3′-adenine and the O2 atom of a 3′-thymine forming a cross-strand hydrogen bond, and the distance from the O2 of its Watson–Crick paired thymine are also indicated. Fig. 3(*c*) shows a similar diagram for a GA·TC step. In all these cases, the cross-strand H2···O2 distances are shorter than the distances in the A·T Watson–Crick base pairs as well as those in the fibre model (Fig. 1*c*).

The histograms in Figs. 2(*c*) and 2(*d*) show the distribution of cross-strand distances in AA·TT steps; the H2···O2 distances are seen to cluster between 2.6–2.7 Å, while a broad maximum occurs between 3.1 and 3.5 Å for the C2···O2 distance for AA·TT steps in DNA oligomers and in protein-complexed structures. These maxima occur at values which are smaller than both the corresponding mean values of the full data set and the distances in the B-DNA fibre model (3.3 and 4.1 Å, respectively). It is also seen from Fig. 2 and Table 2 that the C2···O2 and H2···O2 distances in the Watson–Crick base pairs are generally longer than for the corresponding cross-strand hydrogen bonds, although both are smaller than the fibre-model values. Both sets of C2–H2···O2 angles have similar values of about 120° (Tables 2 and 3), while the H2···O2—C2 angle also lies between 110 and 130° for both interactions. Very similar parameter distributions are seen for the (Ade)C2—H2···O2(Cyt) cross-strand interactions in the GA·TC steps, as evident from Fig. 3(*c*) and Table 3. The slightly longer H2···O2 distances in the Watson–Crick base pairs may be a natural consequence of the base-pair geometry, wherein the H2 atom of adenine cannot come closer to the O2 atom of thymine without the other hydrogen bonds in the A·T base pair becoming distorted. On the other hand, relative movements of the two base pairs in the AA·TT or GA·TC steps can occur quite readily to facilitate shorter H2···O2 cross-strand distances between the neighbouring base pairs.

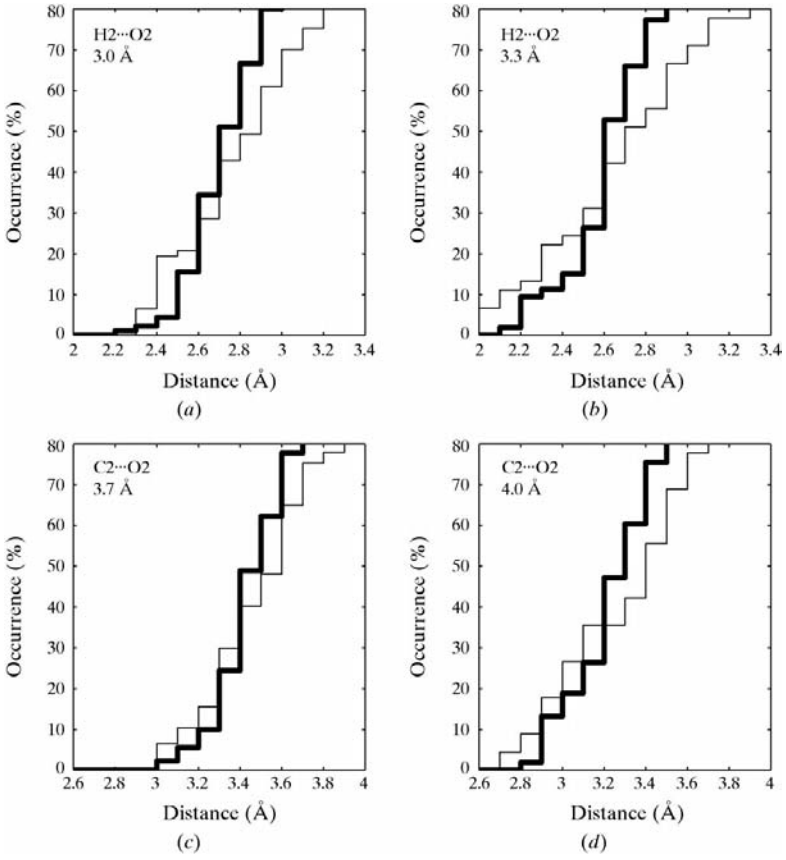

Figure 4
Percentages of H2···O2 and C2···O2 distances in AA·TT steps which are less than a particular cut-off distance are shown as ladder plots (the *x* axis indicates the cut-off value of the step). In each figure, the thick line corresponds to the oligomer data and the thin line to the protein-complexed DNA data set. (*a*) H2···O2 distances in Watson–Crick base pairs and (*b*) H2···O2 distances for the cross-strand hydrogen bond in the minor groove of AA·TT steps. (*c*) and (*d*) show the C2···O2 distances corresponding to (*a*) and (*b*), respectively. For example, 66% of AA·TT steps have H2(3′-A)···O2(3′-T) distance <2.8 Å, while 75% of cross-strand C2···O2 distances are <3.5 Å. The distances corresponding to the fibre-model structure (Chandrasekaran & Arnott, 1996) are given in each figure in order to highlight the considerable reduction in both Watson–Crick and cross-strand H2···O2 and C2···O2 distances in most of the AA·TT steps in the crystals compared with the fibre model.

3.3. Statistical distribution of C—H···O distances in AA·TT and GA·TC steps

In order to draw general conclusions about the geometrical features of C—H···O interactions in the DNA crystal structures, we carried out an analysis of the percentage of H2···O2 and C2···O2 distances (shown in Fig. 2 as fractional frequencies of occurrence) which are shorter than any chosen cut-off distance. Plots of this type are shown in Figs. 4 and 5 for the AA·TT and GA·TC steps, respectively. It is clear that a majority of H2···O2 cross-strand distances in AA and GA steps in the two data sets are shorter than their equilibrium separation distance of 2.86 Å (Cornell *et al.*, 1995). For example, while 51% of H2···O2 distances in the Watson–Crick base pairs are <2.8 Å, 66% of such cross-strand distances meet the same criteria for AA·TT steps in oligomer structures, while the corresponding numbers for the protein-complexed data set are 43 and 51%, respectively. Similarly, 75% of cross-strand C2···O2 distances are <3.5 Å, while 49% of the Watson–Crick base pairs occur within this distance. In the case of GA·TC steps (Fig. 5), a slightly smaller but still significant number (30–40%) of H2···O2 distances are <2.8 Å, while about 40–50% of C2···O2 distances are <3.5 Å in both data sets.

A similar comparison between the minor-groove and major-groove cross-strand interactions in AA·TT steps shows that a considerably larger number of H2···O2 distances are shorter than the cross-strand distance between the H62 atom of 5′-end adenine and the O4 atom of 5′-end thymine on the major-groove side. Thus, while 66% of the H2···O2 distances in AA steps are shorter than 2.8 Å, only about 13% of the (Ade)H62···O4(Thy) distances occur within this cut-off distance. Even a cut-off distance of 3.0 Å gives only a 45% frequency of occurrence for the (Ade)H62···O4(Thy) distance, while it is 81% for the H2···O2 distance in DNA oligomers for AA·TT steps and 60% for GA·TC steps, indicating that cross-strand C—H···O interactions in the DNA minor groove are more common than the cross-strand N—H···O interactions in the major groove (Bansal & Ghosh, 1999).

Cross-strand C2—H2···O2 hydrogen bonds are also found in AA·UU steps, in a dodecamer structure (PDB code 271d) and in inosine-containing steps IA·TC in a similar dodecamer (PDB code

1d77; Xuan & Weber, 1992), and in the IIICCC·IIICCC tracts of a decamer crystal structure (PDB code 286d; Shatzky-Schwartz *et al.*, 1997), suggesting that such cross-strand hydrogen bonds may be a general feature of all dinucleotide steps containing a base with a potential C—H donor group and another with a carbonyl O atom as an acceptor in the minor groove of DNA double helices.

3.4. C—H···N hydrogen bonds

A few of the CA·TG steps also show close cross-strand contacts between the C2—H2 of adenine and the 2-amino N atom of guanine, with H2···N2 distances <2.8 Å and C—H2···N2 angles ≃110°, indicating the presence of C—H···N hydrogen bonds. A representative example of a CA·TG step with cross-strand (Ade)H2···N2(Gua) distance of 2.7 Å is shown in stereo in Fig. 6. The CA·TG steps with short cross-strand H2···N2 distances invariably correspond to CA·TG steps with large twist, negative roll and positive slide, which is associated with a B_{II} backbone geometry (Nagaich *et al.*, 1994). It is interesting to note that these geometries are generally found in structures with CAA fragments, in which the neighbouring AA steps take up an unusual geometry with large positive roll and small twist (for example, in the structures with PDB codes 5dnb, 158d, 1d61, 1d65 and 307d). These are the only AA steps characterized by large cross-strand H2···O2 distances (>3.8 Å), as mentioned in §3.2 and shown in Table 2.

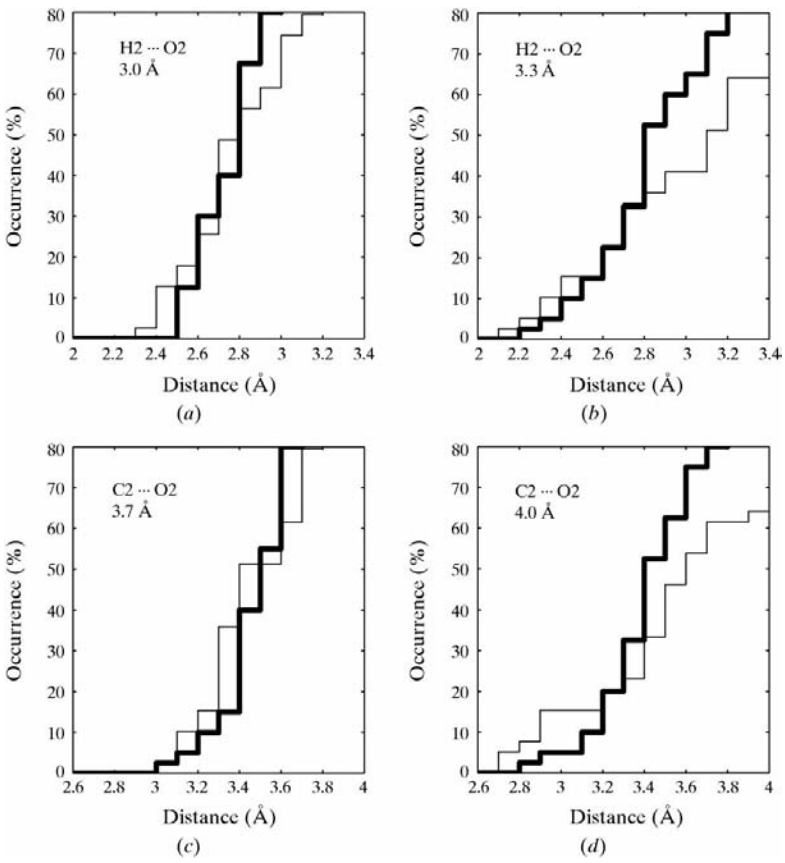

Figure 5
Percentages of H2···O2 and C2···O2 distances in GA·TC steps which are less than a particular cut-off distance are shown as ladder plots (as in Fig. 4 for the AA·TT steps). (*a*) H2···O2 distances in A·T Watson–Crick base pairs and (*b*) H2···O2 distances for the C2H2(Ade)···O2(Cyt) cross-strand hydrogen bond in the minor groove of GA·TC steps. (*c*) and (*d*) show the C2···O2 distances corresponding to (*a*) and (*b*), respectively.

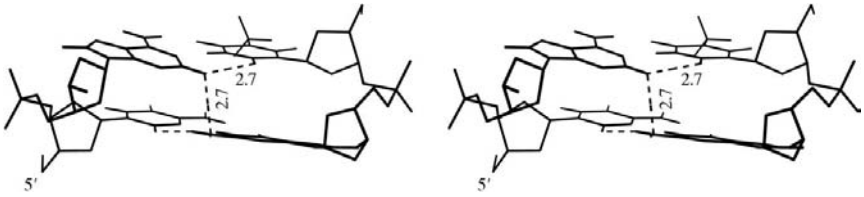

Figure 6
Stereo diagram showing a CA·TG step in a crystal structure (PDB code 5dnb). The H2···O2 distance in the A·T base pair and the cross-strand (Ade)H2···N2(Gua) distance (corresponding to a potential C—H···N hydrogen bond) in the minor groove are indicated.

4. Conclusions

The observed close contacts between the C2–H2 group of adenine and the O2 atom of thymine or cytosine in the AA·TT and GA·TC steps of duplex DNA structures are not surprising, since the C2 atom of adenine is highly electropositive (Cornell *et al.*, 1995), being located between the two electronegative N atoms N1 and N3. This would lead to a favourable electrostatic interaction with an electronegative O2 of the thymine of its Watson–Crick pair and also with that of a 3′ neighbouring thymine or cytosine on the opposite strand. Since the H2 atom also has a small positive residual charge (Cornell *et al.*, 1995), the electrostatic interaction will become even more favourable. However, the local geometry of the AA·TT and GA·TC steps apparently undergoes intrinsic modifications (such as a smaller opening angle of the A·T base pairs and a small negative value for the translation parameter 'slide' compared with a small positive value in the fibre model) so as to bring these groups closer together than in the uniform fibre model. The H2···O2 cross-strand distances are also less than the sum of the van der Waals radii (2.7 Å) in more than 50% of the AA·TT steps and the C—H···O angles are ≥120°, indicating that these interactions are not merely electrostatic in nature but can be assigned as genuine three-centre C—H···O hydrogen bonds, even though they do not fit the standard definition of a hydrogen bond, which would require the C2 atom to be more electronegative than the H2 atom (Steiner & Saenger, 1993).

These results also suggest that the charge assignment for the C2—H2 group in the AMBER 5.0 force field (Cornell *et al.*, 1995) probably needs refinement, since the base-pair energy calculated for an A·T base pair using this force field is about 1.2 kcal mol^{-1} (1 kcal mol^{-1} = 4.184 kJ mol^{-1}) less than the value obtained by the same group using *ab initio* calculations, while the two values are within 0.6 kcal mol^{-1} for the G·C base pair (Gould & Kollman, 1994). If one includes the C—H···O hydrogen-bond energy, which is generally considered to be in the range 1–2 kcal mol^{-1} (Desiraju, 1991), then the agreement for the A·T base-pair energy is substantially improved. It has been shown that C—H···O interactions in alkanes reverse the normal $C^+—H^-$ bond polarization (Wiberg *et al.*, 1991) and it is likely that a similar effect could occur at adenine C2—H2 groups in the vicinity of an electronegative carbonyl O atom, particularly in the minor groove of AA·TT or GA·TC steps, where two equally strong C—H···O interactions seem to be occurring simultaneously. Observation of C—H···O interactions during the molecular-dynamics simulations of RNA has also led to the conclusion that the parameters for these weak interactions need to be refined in the various force fields (Auffinger & Westhof, 1998).

The C2—H2···O2 hydrogen bonds within the A·T base pairs could be a natural consequence of Watson–Crick pairing; however, the favourable minor-groove cross-strand interactions arise as a consequence of the specific sequence-dependent geometry of the AA·TT and GA·TC steps. The propeller twist alone does not lead to close C—H···O contacts in the minor groove of the AA·TT and GA·TC steps and the latter also does not have exocyclic groups capable of favourable cross-strand interactions in the major groove [in fact, the close interaction between the (Gua)O6 and (Thy)O4 groups will be strongly repulsive]. Hence, it appears that in addition to tightly bound water molecules, the C—H···O cross-strand interactions in the minor groove could also play an important role in determining the structural invariability of AA·TT and GA·TC steps, as well as A_n·T_n and GA_n·T_nC tracts in DNA oligomers. This local geometry of AA·TT steps gives rise to a narrow minor groove when such sequences occur in consecutive steps (A-tracts) and also imparts rigidity to the poly(dA)·poly(dT) structure.

This work was partially supported by financial assistance from the Department of Biotechnology, India.

References

Auffinger, P. & Westhof, E. (1998). *Curr. Opin. Struct. Biol.* **8**, 227–236.

Bansal, M., Bhattacharyya, D. & Ravi, B. (1995). *Comput. Appl. Biosci.* **11**, 281–287.

Bansal, M. & Ghosh, A. (1999). *Perspectives in Structural Biology*, edited by M. Vijayan & N. Yathindra, pp. 353–364. Hyderabad: Universities Press.

Bella, J. & Berman, H. M. (1996). *J. Mol. Biol.* **264**, 734–742.

Chakrabarti, P. & Chakrabarti, S. (1998). *J. Mol. Biol.* **284**, 867–873.

Chandrasekaran, R. & Arnott, S. (1996). *J. Biomol. Struct. Dyn.* **13**, 1015–1027.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179–5197.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.

Desiraju, G. R. (1996). *Acc. Chem. Res.* **29**, 441–449.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S. & Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J. & Zhurkin, V. B. (1989). *J. Mol. Biol.* **205**, 787–791.

DiGabriele, A. D. & Steitz, T. A. (1993). *J. Mol. Biol.* **231**, 1024–1039.

Drew, H. & Dickerson, R. E. (1981). *J. Mol. Biol.* **151**, 535–556.

Fabiola, G. F., Krishnaswamy, S., Nagarajan, V. & Pattabhi, V. (1997). *Acta Cryst.* D**53**, 316–320.

Gould, I. R. & Kollman, P. A. (1994). *J. Am. Chem. Soc.* **116**, 2493–2499.

Leonard, G. A., McAuley-Hecht, K., Brown, T. & Hunter, W. N. (1995). *Acta Cryst.* D**51**, 136–139.

Luisi, B., Orozco, M., Sponer, J., Luque, F. J. & Shakked, Z. (1998). *J. Mol. Biol.* **279**, 1123–1136.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L. & Zhurkin, V. B. (1998). *J. Mol. Biol.* **277**, 1129–1140.

Nagaich, A. K., Bhattacharyya, D., Brahmachari, S. K. & Bansal, M. (1994). *J. Biol. Chem.* **269**, 7824–7833.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). *Nature (London)*, **330**, 221–226.

Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F. & Shakked, Z. (1997). *J. Mol. Biol.* **267**, 595–623.

Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). *Biochemistry*, **37**, 8341–8355.

Shui, X., Sines, C. S., McFail-Isom, L., Van Derveer, D. & Williams, L. D. (1998). *Biochemistry*, **37**, 16877–16887.

Steiner, T. & Saenger, W. (1993). *J. Am. Chem. Soc.* **115**, 4540–4547.

Taylor, R. & Kennard, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.

Tereshko, V., Minasov, G. & Egli, M. (1999). *J. Am. Chem. Soc.* **121**, 3590–3595.

Wahl, M. C. & Sundaralingam, M. (1997). *Trends Biochem. Sci.* **22**, 97–102.

Wiberg, K. B., Waldron, R. F., Schulte, G. & Saunders, M. (1991). *J. Am. Chem. Soc.* **113**, 971–977.

Xuan, J.-C. & Weber, I. T. (1992). *Nucleic Acids Res.* **20**, 5457–5464.